British Journal of Social Work (2026) **00**, 1–21
https://doi.org/10.1093/bjsw/bcag017

Critical Commentary

Just because we can does not mean we should: Responsible and meaningful uses of administrative data in the evaluation of complex interventions in children's social care

Lisa Holmes [1,*], Rick Hood [2], Louise Mc Grath-Lone [3], Rocio Méndez Pineda [3], Jeongeun Park [1], Jonathan Scourfield [4], Joanna Soraghan [5], and Jenny Woodman [3]

[1]School of Education and Social Work, Essex House, University of Sussex, Essex House, Brighton, BN1 9QU, England
[2]Department of Social Work and Social Care, Kingston University, Penrhyn Road campus, Kingston University, Kingston-upon-Thames, KT1 2EE, England
[3]Social Research Institute, University College London (UCL), Woburn Aq, London, WC1H 0AA, England
[4]CASCADE, Spark, Cardiff University, Maindy Road, Cardiff, CF24 4HQ, Wales
[5]CELCIS, University of Strathclyde, Cathedral St, Glasgow, G4 0LG, Scotland

*Corresponding author. School of Education and Social Work, University of Sussex, Essex House, Brighton, BN1 9QU. E-mail: l.j.holmes@sussex.ac.uk.

Abstract

Administrative data are increasingly used in impact evaluations of complex interventions in children's social care (CSC), offering advantages such as longitudinal, population-level coverage, and reduced burden on families and practitioners. However, administrative data alone are insufficient to support meaningful and responsible evaluation. Drawing on

insights from two webinars hosted by the Children's Social Care Data User Group, which focused on presentations about uses of administrative data for evaluations, we identify three core challenges that administrative data cannot resolve in isolation: understanding the intervention, creating a valid comparison group, and measuring relevant outcomes. We explore how the absence of detailed programme theory, limited information on intervention implementation, and lack of nuanced outcome measures constrain the interpretability of evaluation findings. We highlight the risks of confounding in quasi-experiments when key contextual variables are missing from administrative datasets. While we acknowledge the value of administrative data, we caution against over-reliance on datasets designed for operational rather than evaluative purposes. We offer recommendations for more robust evaluation practice, including commitments to building programme theory, prospective data collection, capturing lived experiences, and transparent reporting of study limitations. These recommendations align with emerging policy frameworks and evaluation strategies but require adequate time, funding, and interdisciplinary collaboration.

Keywords: administrative data; children's social care; complex interventions; evaluation.

Accepted: January 2026

Introduction

Impact evaluations carry significant weight in shaping policies and practice, so it is essential they are conducted ethically and responsibly and can tell us something useful. This critical commentary examines the increasing use of administrative data for impact evaluations in the field of children's social care (CSC) in the UK. This is within the context of growing attention to increases in the number of children and families referred to CSC and recent recognition that one in four children will have needed support from CSC services by the time they are eighteen (Jay et al., 2025). Relevant interventions include manualized programmes for individuals and families, practice frameworks or models for professionals, models of decision-making and ways of structuring services.

Within the UK, CSC includes children's centres, early help, family support, child protection and provision of out-of-home care, mainly provided by municipal governments or 'local authorities' with varying size, population and ways of delivering their services. Although terminology varies across the four UK nations, we use CSC throughout this paper for simplicity. We also consider 'CSC-relevant interventions', which are developed to work alongside CSC services or to divert children and families from CSC (La Valle et al., 2019).

Both CSC and CSC-relevant interventions tend to be highly 'complex' in the degree of local adaptation (flexibility), the number of interacting components involved in the intervention, their interaction with CSC and the wider system and the range and nature of the intended behaviour or

systems change (Skivington et al., 2021; Caffrey and Browne 2023). In many of these interventions, there is an emphasis on 'how' things are done with families as well as 'what' is done, which makes them hard to evaluate (Gillingham 2018; James and Holmes 2024).

In recent years, an increasing number of CSC-relevant impact evaluations have reported no evidence that the intervention improves outcomes for children and families. These include impact evaluations of integration of health and social care services for children in Scotland (Ottaway et al., 2023), targeted health visiting (home visiting for preschool children) (Anderson et al., 2023; Cavallaro et al., 2024) and Special Educational Needs and Disability (SEND) services in England (Gilbert et al., 2025) and the Social Workers in School programme (Westlake et al., 2024). These results are difficult to interpret. Do the results tell us that the intervention does not help children, young people and families at all, or only that no change was detected in specific outcome measures? Are the study methods such that we could not detect any impact even if it existed?

These 'no evidence of effect' impact evaluations all rely primarily on administrative data. These are the data routinely captured about a child, young person or family and their contact with CSC and other public services (see Box 1). There are many benefits to using administrative data and the authors of this paper are all advocates for using administrative data in CSC-relevant research. One significant advantage of administrative data is that every child referred to CSC will be captured within national (and local) administrative data (Box 1). This means we do not have to rely on information given by families, who tend to under-report contact with CSC (Zhang et al., 2020).The benefits of administrative data have led to government investment in advancing its use in research. For example, there has been investment in Administrative Data Research (ADR) UK, Foundations (the What Works Centre for Children and Families) and specific grants such as the 'Community Catalyst' (Administrative Data Research UK 2025). There have also been concerted efforts to link administrative datasets and make these available to researchers and analysts, via initiatives such as the SAIL databank. These initiatives have facilitated useful longitudinal analyses, and have helped us to better understand the prevalence of CSC involvement, and involvement with other agencies (through linked datasets) (Hodges et al., 2025). However, administrative datasets are collected for operational (as opposed to research) purposes, and there are intrinsic limitations around data quality and the type and range of available information. There also are limitations associated with missing data, or mis-recorded information, and inconsistencies in recording and reporting (data quality) (Harron et al., 2017; Jones et al., 2018; Fitzsimons and McCracken 2020). Finally, there are also significant challenges around data acquisition pathways, approvals, controls on access and disclosure, metadata (i.e. documented descriptions of data meaning and/or solutions to tricky data issues), high researcher skill needed and create and destroy

Box 1. National CSC-related administrative datasets within England.
There are similar data sets in the other UK nations to those from England, some of which have more advanced linkage to health data (e.g. primary care data) and already have household linkages to investigate clustered experiences and service use within families.

Children in Need Census: Data related to children referred to and assessed by children's social care. https://www.gov.uk/government/collections/statistics-children-in-need

Children Looked After: Data about children under local authority care, including adoption. https://www.gov.uk/government/collections/statistics-looked-after-children

National Pupil Database: A national database that brings together information about the educational experiences of children in state school in England. https://www.find-npd-data.education.gov.uk/categories

Education and Child Health Insights from Linked Data (ECHILD): A national data resource linking together data from hospitals, schools, and social care services for children in England. https://www.echild.ac.uk/

As outlined in their 2025 Evaluation Strategy (Department for Education 2025), the Department for Education (DfE) is also working to link data from children's social care and education to data from the Department for Work and Pensions (DWP) on benefits and employment, from HM Revenue and Customs (HMRC) on income, from Ministry of Justice (MoJ) on contacts with the criminal justice system and from demographic and detailed household information from the Office for National Statistics (ONS) census.

models which may lead to duplication of effort (Jones et al., 2018). In short, there are both great benefits and great challenges to the use of administrative data for research which have been well documented.

This critical commentary grew from two webinars hosted by the Children's Social Care Data User Group (CSCDUG) in 2024, prompted by a growing awareness of the 'no evidence of effect' evidence base. The first of these webinars focused on uses of national administrative data for evaluation and the second focused on uses of linked local data (including administrative data). Created in 2018 by researchers working in the field, the CSCDUG provides a forum to share expertise and learning between users of CSC administrative data across academia, policy and practice in the UK. The CSCDUG now has around 200 members, comprised of academics, local authority children's services and national government analysts, policy makers, and representatives from national charities working in the field of CSC. All CSCDUG members were invited to join the authorship team for this critical commentary. The requirement for inclusion was a commitment to participate in an online meeting, one in-person workshop and contribute to the writing. This critical commentary was conceptualized and prepared by eight academics from a range of disciplines, predominantly from social work, social policy, public health, and epidemiology, all with direct experience of analysing administrative datasets and conducting CSC evaluations. We structure our commentary into three challenges of impact evaluations that cannot be solved by routinely collected

administrative data alone: understanding the intervention, creating a comparison group, and measuring relevant outcomes. In Table 1, we provide a summary of the main interventions, and associated evaluations that we refer to in the text, including links to evaluation reports and published outputs. We highlight responsible approaches to these challenges which, we argue, are essential to generate a usable evidence base to inform CSC policy and practice.

Challenge 1: describing and understanding CSC-relevant interventions

For an evaluation to tell us anything useful about whether interventions are effective, we first need to know what we are evaluating, both in theory (to design the evaluation) and in practice (to interpret the results). This information is not routinely captured in administrative data (Sebba et al., 2017; Fitzsimons and McCracken 2020).

The intervention in theory

Programme theory, sometimes called a theory of change, summarizes how the intervention works, i.e. how all its elements and activities are designed to lead to intended outcomes (Skivington et al., 2021). Knowing the programme theory helps us answer necessary questions for planning evaluations: what is the ultimate goal of the intervention, which children and families do we anticipate to benefit, where might we see early steps towards this ultimate goal, and should we be analysing results separately for certain sub-groups of children and families or for specific intervention sites because we would expect the intervention to work differently in these contexts?

Well described programme theory is particularly important for complex interventions which are adapted to local contexts and variably implemented across sites. Programme theory can identify the core 'functions' of a complex intervention (the essence of the intervention without which we cannot expect change to take place) which should remain in place even if the activities and components are modified for a local context (Hawe et al., 2004; Evans et al., 2021). This approach recognizes that some adaptation across implementation sites may be necessary and beneficial, while identifying that some 'essential elements' must be preserved.

An evaluation of Multiagency Safeguarding Hubs (MASH) and a study on the integration of CSC and health services in Scotland found that a review of the literature and other publicly available material on the intervention was on its own insufficient to develop detailed programme theory that could adequately support the design and interpretation of an

Table 1. Descriptions of CSC interventions referred to in the text, with details of impact evaluations.

Intervention	Key impact evaluations (including details of administrative data components)
Signs of Safety (SoS): Trade-marked strengths-based, safety-orientated approach to CSC casework drawing on Focused Brief Therapy: with toolkit for family participation and choice and philosophical underpinnings about how to work with families (Sheehan et al., 2018; Baginsky et al., 2019; Caffrey and Browne 2022, 2023). 86% ($n = 128$) of local authorities in England reported using SoS in 2017 (Baginsky et al., 2019). High variation in implementation with a 'pick and mix' approach (Sheenan et al., 2018) Theorized through Realist[a] evaluation (Baginsky et al., 2019) and syntheses (Sheenan et al., 2018; Caffrey and Browne 2022, 2023).	No impact evaluations in the UK using administrative data. Realist evaluation using qualitative methods in 10 sites found staff and families reported positive effects (Baginsky et al., 2019). An Australian impact evaluation using administrative data and retrospective quasi-experimental[b] methods found increases in children in care and substantiated notifications after SoS was introduced, which the authors highlighted as difficult to interpret (Baginsky et al., 2019).
Family Group Conferencing (FGC): A form of family group decision making based on partnership working with a child's wider family network (Family Rights Group 2025). Proposed legislative change for England would require all local authorities to offer Family Group Decision Making where children are at risk of entering the care system (Children's Wellbeing and Schools Bill 2025). There are many variants of FGC as implemented, though core principles exist (Bredewold and Tonkens 2021). 79% ($n = 124$) of local authorities in the UK were using FGC in 2022 but with variability in implementation (Wood et al., 2022). Theorized through Realist[a] reviews and evaluation (Stabler et al., 2019, 2025; Scourfield et al., 2024; Bernheim et al., 2025).	Internationally, there are many impact evaluations of FGC, many of which rely on administrative data, with contradictory results (Nurmatov et al., 2020), although the only UK RCT, using administrative data, found a positive result for reduction in care entry at pre-proceedings stage (Taylor et al., 2024). There is an in-progress evaluation of FGC, which includes a study of implementation, development of programme theory, evaluation of self-reported family outcomes and a quasi-experimental study using administrative data[b] (Scourfield et al., 2024).
Multiagency Safeguarding Hubs (MASH): A service model at the front door of CSC to facilitate multi-agency information sharing and decision making between social workers, health professionals, police, education and others (Home Office 2014; Méndez Pineda et al., 2025). 93% of 'front door' services reported having a MASH in 2023 ($N = 99/106$) with very high variation (Méndez Pineda et al., 2025). Beginning to be theorized (Méndez Pineda et al., 2025).	An early impact evaluation of MASH in London (before/after intervention design) used administrative data and found reductions in times from referral to decision but more children referred to MASH for abuse and neglect after implementation (Crockett et al., 2013). The authors concluded it was difficult to interpret the findings on referrals. There is an impact evaluation in-progress as a PhD study, using retrospective quasi-experimental[b] methods to ascertain the impact of MASH on re-referrals and escalation of cases into child protection (Méndez Pineda 2026).
Lifelong Links: An approach developed by Family Rights Group to build and maintain supportive relationships and networks for children in care and care leavers, via a trained independent Lifelong Links coordinator who works with the child.	The evaluation of the trial in 12 local authorities in England used a retrospective quasi-experimental design[b] and administrative data and found higher placement stability of children who received Lifelong Links (Holmes et al., 2020).

(continued)

Table 1. (continued)

Intervention	Key impact evaluations (including details of administrative data components)
Lifelong Links was initially trialled in 12 local authorities in England with funding from the Department for Education CSC Innovation Programme (Holmes et al., 2020). Lifelong Links was theorized based on the learning from the (Care Inquiry 2013) and the under-pinning principles of Family Group Conferences.	
The structural integration of adult and children's health and social care services within newly established Health & Social Care Partnerships (HSCPs) in Scotland comprises bringing together services (Anderson et al., 2023). The rationale for structural integration was presented in the Scottish Government's Health and Social Care delivery plan (Scottish 2016) and underpinned by the findings of the Commission on the Future Delivery of Public Services (Christie 2011) but there has been no research to produce programme theory.	Anderson et al. (2023) used a retrospective quasi-experimental design and 25 area-level indicators in administrative data across a range of domains (e.g. child protection, education, health). The study did not find consistent evidence of changes in outcomes that could be attributed to the extent of integration of children's services but concluded that this may be due to methodological challenges.

[a]Realist evaluation is an approach that develops theories about how interventions work by asking 'what works for whom, in what contexts, in what respects, and how?' (Pawson and Tilley 1997). This approach assumes that whenever a programme is implemented, it tests a theory about what might cause change, even if that theory is not explicit. Realist evaluation makes these theories explicit.
[b]Quasi-experimental designs test causal hypotheses about programme effectiveness but do not use the random assignment to treatment and control groups that characterizes randomized controlled trials (Shadish and Luellen 2006). The key challenge in quasi-experimental designs is demonstrating that any observed differences in outcomes between treatment and comparison groups are due to the intervention rather than to systematic differences between the groups (confounding).

evaluation (Ottaway et al., 2023; Méndez Pineda et al., 2025). In other words, these authors concluded that qualitative data collection was necessary to plan a robust impact evaluation.

There are some very strong examples of programme theory underpinning evaluation design in CSC. One is the evaluation of Family Group Conferencing for children and families, which was clear about the underlying programme theory (see Table 1) However, there are also examples of evaluations using administrative data that have been funded without explicit theoretical underpinning. For example, the team conducting the MASH impact evaluation (see Table 1) faced uncertainty about which were the most appropriate outcomes to measure for which groups of children in the evaluation, which was in the end addressed by separately funded programme theory research (Méndez Pineda et al., 2025).

The intervention as implemented

We also need to know how the intervention has been put into practice on the ground, which is no small task for complex interventions. By design, the activities of a complex intervention vary from place to place and with adaptation to suit local contexts. Variation in complex interventions may also arise from, e.g., differences in timescales, eligibility decisions, legal frameworks, and how—and indeed whether—different aspects of the intervention are implemented across time or intervention sites (Holmes 2021; Anderson et al., 2023; Crawford et al., 2025). Research about Signs of Safety has found that use of the tool kit was easier for social workers than adopting the philosophical underpinnings of the intervention and Family Network meetings could be underused (Table 1). The way of working with families, not the tool kit, constitutes the core 'functions' of this intervention, on which it relies to work.

This variation may lead to an overall evidence-base which is apparently contradictory, as Stabler and colleagues argue is the case for Family Group Conferencing (Stabler et al., 2025, Table 1). To make sense of a complicated evidence-base, we need to know which version of an intervention was implemented and the extent to which the intervention was delivered according to its underlying theory and design. In other words, we need to know how far 'functional fidelity' has been achieved (Evans et al., 2021). If functional fidelity is not achieved, we should not be surprised if the intervention does not seem to work on this occasion.

Business-as-usual services

It is also essential to describe what services and support are being delivered in the non-intervention groups, to help explain why there appears to be or not be a difference in outcomes between these groups. It is possible that robust business-as-usual services already look similar to the intervention: Signs of Safety, for example, has been described as 'just good social work' (see Table 1; Baginsky et al., 2019). If 'good social work' operates in the comparison sites, we would not necessarily expect to see any additional benefit from implementing Signs of Safety.

Data recording and collection

When evaluations are built into the roll-out of an intervention, local authorities may agree to collect new information about implementation within their administrative data, including who was offered and received the intervention. For example, in a study of 12 local authorities implementing Lifelong Links, some local authorities placed a 'flag' in young

people's records, or created custom forms, indicating those who were eligible for the intervention, had been offered the intervention, had received or declined of intervention (with reason for declining) and returned this data to the research team for matching with administrative data and analysis (Holmes et al., 2020). Flags such as these are extremely valuable but insufficient for assessing functional fidelity completely, which needs additional information from qualitative research, as was also collected in this evaluation (Holmes et al., 2020). Alternatively, this information can be captured as part of the study, as with the Social Workers in Schools trial (Westlake et al., 2024). A range of different methods and types of data provide crucial context for interpreting outcomes (Ginsburg et al., 2021). Arranging for local authorities to collect extra data within their administrative systems or setting up observations or a survey of current practice requires dedicated funding and planning in 'real-time' (i.e. as an intervention is being delivered or rolled out). This is known as 'prospective' data collection.

When real-time individual level data about the intervention is not available, researchers have used creative approaches. For example, an evaluation of integration of CSC and health services in Scotland relied on publicly available information indicating that an intervention was taking place in a local authority–see Table 1 (Anderson et al., 2023; Ottaway et al., 2023). An evaluation of MASH used a survey about previous practice and Ofsted reports to capture start date and intervention characteristics (Méndez Pineda et al., 2025). There are problems with these approaches. Using area level data on interventions can make it more challenging to detect any positive impact that may exist as we can only measure potential impact in *all* children and families in the area, regardless of whether they received the intervention or not. Additionally, publicly available data may only capture one aspect of the intervention as implemented, which may not be the most important element, making it difficult to ascertain if key parts of the intervention took place (see Table 1). Surveys about previous practice are subject to recall bias: staff cannot always remember what was happening several years ago or there may have been staff turnover. The problems with retrospectively ascertaining whether and how an intervention took place leads to studies that are less able to detect any positive impacts of an intervention if they exist. Evaluation planning should therefore take place before the rollout of an intervention, providing the opportunity for prospective data capture that will explain key information on implementation.

Challenge 2: comparison groups

An impact evaluation needs an appropriate comparison group to estimate outcomes for those who did **not** receive the intervention as well as those who **did**. This comparison group needs to be similar to the intervention

group in all relevant characteristics before the intervention begins (baseline equivalence). If the comparison and intervention group are not similar at baseline, we cannot rule out the possibility that any subsequent differences in outcomes between the two groups are a result of underlying differences between the groups rather than a result of the intervention ('confounding by indication') (Tucker 2022; Hernán and Robins 2024). This confounding will make it more challenging to detect any positive impact of the intervention if it exists and/or may even make it look like an intervention caused worse outcomes, when that is not necessarily the case.

For example, in an impact evaluation of service-as-usual health visiting in England, infants who received 'enhanced', more frequent, health visiting showed higher rates of unplanned hospital use (Woodman et al., 2025; Bunting et al., 2026). This finding was counter to the team's original hypothesis: increased contact with the health visiting team in the early weeks of a child's life could build the capacity of families to prevent and manage common childhood health problems. The findings are difficult to interpret, as the authors highlight (Bunting et al., 2026). It could mean that more intensive health visiting increased serious illness and injury among infants leading to more unplanned hospital use. Or, increased hospital use may represent appropriate health-seeking behaviour from families, perhaps encouraged by health visiting teams. A third explanation is that there were important differences between the intervention and control groups at baseline which meant that the intervention group was both more likely to receive extra health visiting and more likely to have unplanned hospital use for their baby (Woodman et al., 2025; Bunting et al., 2026). Qualitative work suggested that practitioners were using a range of factors, including their assessment of the physical home environment, family structure, and social isolation to decide whether a family needed more health visiting services, all factors which could also influence a baby's likelihood of having unplanned hospital use (Woodman et al., 2025; Bunting et al., 2026). None of this information was captured in routine collected administrative data and therefore could not be used for 'matching' the families who did and did not have enhanced health visiting (Woodman et al., 2025; Bunting et al., 2026). If qualitative investigation of important 'matching' factors had been conducted at the planning stage of the impact evaluation (rather than afterwards), the authors might have concluded that they could not rely solely on retrospective and routine administrative data for this impact evaluation.

In CSC, any practitioner decisions about who to put forward or prioritize for any interventions and family decisions about whether to take-up or decline child welfare interventions will be influenced by a complex decision-making ecology (Hood 2022), including the family's social circumstances and inter-family relationships; as well as local system factors, such as local practice cultures of thresholds and risk-aversity. Most of

these factors will not be captured in the coded administrative data. This means that, like in the health visiting study above, relying on coded administrative data to create a 'matched' comparison group becomes problematic. Randomization is the surest way to make sure that the intervention and comparison groups are similar in important characteristics. However, this is not always possible or acceptable. In these scenarios, researchers often adopt 'quasi-experimental' designs to exploit naturally occurring variation across time, locations or intervention coverage to generate a comparison group (Tucker 2022). One commonly used quasi-experimental technique is statistical 'matching', which creates groups of people based on the probability of having similar characteristics who did and did not receive the intervention (Heinrich et al., 2010). This approach relies on having the correct information for those that did and did not receive the intervention. It is challenging to adequately match groups when relying solely on information within CSC administrative data, as necessary information on participant characteristics, or family environment is not routinely captured.

Challenge 3: defining and measuring outcomes

The complexities of capturing the 'right' outcomes associated with CSC are well documented (Forrester 2017; La Valle et al., 2019). The programme theory of an intervention assists researchers in deciding which outcomes to measure for which children or families. For example, the programme theory for SoS suggests that an impact evaluation might consider contact time between families and social workers, changes in social workers' scaling of 'worry' about a child, or parental empowerment and satisfaction as outcomes–see Table 1 (Sheenan et al., 2018; Nurmatov 2020). Programme theory of MASH suggests that if we want to know whether certain models of MASH are working, we should measure repeat referrals into CSC, earlier receipt and higher volume of early help and preventive services for families, practitioners' confidence and assurance in their decision-making and families' views of the service (Méndez Pineda et al., 2025).

In these examples, few of the relevant outcomes are available in routinely collected administrative data, as these data predominantly capture service use (Hood 2019; La Valle et al., 2019; Hood et al., 2024). For example, administrative data from CSC can provide information about changes in the rates of children in care, children in need or re-referrals to CSC services (Anderson et al., 2023; Ottaway et al., 2023). However, it contains limited information about practitioner concerns about a child beyond key 'factors for assessment' (Hood et al., 2023), and nothing about the quality of family interactions, relationships within or outside the family, family

experiences or views, or the quality of a family's relationships with practitioners and the service.

Unlike in health care and adult social care, users of CSC services are not routinely surveyed about their satisfaction with the services they receive and so this does not form part of the administrative data return (Office for Health Improvement and Disparities 2020). There are examples from the Adult Social Care Outcomes Toolkit that captures the effectiveness of social care interventions (Netten et al., 2012). However, these data are not currently integrated into corresponding administrative datasets for research purposes. Routinely collecting the child's voice in a way that allowed it to be included in analyses of administrative data would align with Article 12 of the United Nations Convention on the Rights of the Child—'every child has the right to express their views, feelings and wishes in all matters affecting them' (UNCRC 1989).

Impact evaluations often make use of demand measures (i.e. rates of referral and service provision) that are readily available in routine administrative datasets. For example, the main outcome measure used in the Social Workers in Schools trial was the rate of 'Section 47' child protection investigations that the intervention was designed to prevent. Such measures were also dominant in the Department for Education's CSC Innovation Programme (Sebba et al., 2017). There will always be debate about the 'right' outcome to select, but demand measures will not always be the most appropriate or logical (Drake and Jonson-Reid 1999; Gubbels et al., 2023). Using these measures as outcomes may lead to impacts not being detected (even if impact does in fact exist) and misinterpretation of findings. Suh and Holmes (2022) have highlighted how this is exacerbated in economic evaluations which emphasize outcomes that are monetizable as well as readily available in administrative datasets.

Linking administrative datasets across CSC, health, education, police and family justice promises a wider range of available outcomes for use in impact evaluations. However, without routine surveys of people's experiences and perceptions, administrative data will always focus primarily on demand and process measures. Even widely linked administrative data will not provide key information about families, such as about their networks, interactions, views and relationships with each other and with services. Additionally, some commonly used CSC-relevant outcomes in health administrative data are ambiguous as outcome measures in evaluations. Studies on enhanced or intensive health visiting have reported higher healthcare contacts for maltreatment or adversity in the intervention group, as measured by coded administrative data. This could be either a positive effect of the intervention indicating appropriate healthcare use and increased recognition, recording and treatment. Or this could be negative effect, indicating increased adversity or increased illness/injury necessitating healthcare (Cavallero et al., 2024; Woodman et al., 2025; Bunting et al., 2026). This outcome will also be shaped by recording practices

within healthcare settings, which will likely differ over time and between local areas. As we explain above, higher rates of adversity-related healthcare use in the enhanced health visiting groups may also signal imperfectly matched intervention and control groups, i.e. residual confounding by indication, see above. There are examples of some study teams moving away from their original plans to use adversity or maltreatment-related healthcare use as a primary outcome due to ambiguity of interpretation in impact evaluations (Woodman et al., 2022).

Linking administrative data to cohort studies may supply a wider range of relevant outcomes beyond contact with services. Cohort studies collect social, health-related and biomedical data directly from large, nationally representative samples of families repeatedly at set time points from soon after birth into late adulthood. Examples include the UK's Millennium Cohort Study (Connelly and Platt 2014) and local cohorts such as Born in Bradford (Wright et al., 2013). Nevertheless, it is important to consider the bias that can arise through the linkage process (Harron et al., 2017) and this must be assessed and accounted for within an evaluation to ensure that accurate conclusions can be drawn. These studies to date in the UK have extremely limited self-report data on CSC and little linkage to administrative data (Zhang et al., 2020). In the future, the intention is to improve such linkage and funding has just been announced to investigate the feasibility of a new birth cohort which will be focused specifically on collecting data for child welfare research (CLS 2017; Swansea University 2024).

Recommendations

Having set out three challenges of impact evaluations that cannot be solved by routinely collected administrative data alone, even if widely linked, we now focus on a set of recommendations for future evaluations. Our recommendations chime with other guidance on meaningful and responsible evaluations (Skivington et al., 2021; Craig et al., 2025).

Commitment to building programme theory as its own endeavour

For evaluations in CSC to be robust and interpretable, they need to be based on programme theory. If there is not already an advanced evidence base detailing programme theory, a primary study will be required in advance of the impact evaluation. Building programme theory of promising CSC-relevant interventions as its own endeavour would create an evidence-base to springboard robust impact evaluations. This is important because impact evaluations are often funded to deliver to a tight timeline to feed into policy decisions.

Commitment to prospective data capture on

1. **Implementation:** To ensure a meaningful and rigorous evaluation, it is essential to capture details of how the intervention was delivered. Ideally, this should include individual-level data on eligibility and receipt of the intervention and site-level information on how the intervention was implemented (including adaptations) and whether the key functions of the intervention were in place (functional fidelity).
2. **Family characteristics and outcomes:** Primary data collection for the impact evaluation would allow research teams to collect necessary information for matching (in non-randomized designs) and capturing the most relevant outcomes according to the programme theory of the intervention. Research teams could also consider whether cohort studies could be linked to administrative data to provide this information without new data collection.

Capturing views and experiences

Collecting the views and experiences of children, families, and practitioners provides valuable context for evaluations, and qualitative insights should be used to inform quantitative analysis. For instance, it is essential that data science and analytical proficiency are complemented with the expertise and contextual knowledge of social work practitioners within an evaluation team. Moreover, it would greatly support both internal and external evaluation if children and families were routinely surveyed and their experiences incorporated into the administrative data returns for CSC. Given government investment in the National Outcomes Framework for CSC, we consider this to be a priority area for reform.

Commitment to robust consideration of study limitations at the planning stage

A commitment to transparency in evaluation design is essential. This includes providing detailed information about the methods used, assumptions made and the limitations of the data. Full consideration should be given *at the planning stage* to how any assumptions or limitations may influence the impact evaluation findings and the conclusions that can be drawn. If study findings are likely to be challenging to interpret, funders and research teams should consider whether it is responsible to proceed with the evaluation.

Conclusion

Administrative data is a powerful resource for research that informs CSC policy and practice, including through impact evaluations of interventions. However, care is needed. For an impact evaluation to have a good chance of producing interpretable results, it should be designed using detailed programme theory which is built from empirical research, and include prospectively collected information on the intervention in practice and the most relevant outcome measures, some of which will very likely need collecting directly from families, practitioners or the service. This means that linking multiple administrative datasets to each other is unlikely to solve the key issues in using these data to evaluate interventions in CSC, although linked cohort study data would offer more promise. The approach we suggest instead relies on skilled research teams working alongside experts in CSC practice and people with lived experience to conduct preliminary work to support planning of the impact evaluation and collect new data over time on the intervention and families as part of the study.

There are examples of where this approach is already being put into practice and where many (but not all) of our suggestions are already in place (Scourfield et al., 2024). Some of our arguments about meaningful and responsible evaluations are reflected in the Department for Education's 2025–9 Evaluation Strategy objectives (Department for Education 2025). The Strategy states that all policies should have a Theory of Change at the business case stage, which aligns with our calls for building programme theory as its own endeavour. The Strategy also states that evaluation plans should be built into the full rollout of a policy after piloting, which would support prospective data capture for evaluation.

We also observe that policy timeframes and funder expectations of research timelines may encourage research teams towards a less thorough approach. Sometimes, there are mixed messages from funders, with an emphasis placed on 'relevant and meaningful' evaluations, but with funding timeframes that do not exceed 18 months. These short (for research) timelines preclude thorough development or testing of programme theory and sufficient prospective data capture for robust impact evaluation. Building of the evidence-base outside (before) policy decisions may mitigate some of the discord between research and policy timelines, i.e. through provision of an evidence-base that policy can draw on, as and when needed.

The government, funders, and researchers all have a role in ensuring that administrative data is used responsibly and meaningfully in impact evaluations of CSC and CSC-relevant interventions: just because we can conduct an impact evaluation at (relative) speed relying solely on administrative data doesn't mean we should.

Conflicts of interest. None declared.

Funding

This study was supported in part by National Institute for Health and Care Research (NIHR) Public Health Research Programme (NIHR129901), including through J.W. and L.M.G.L.'s time. J.P.'s time working on the study was funded by UK Research and Innovation (UKRI) (Medical Research Council) as part of the ReThink programme (MR/W002132/1). The views expressed are those of the author(s) and not necessarily those of the NIHR, the Department of Health and Social Care or UKRI (MRC). J.S. receives funding as a Health and Care Research Wales Senior Research Leader.

References

Administrative Data Research UK. (2025) 'Community Catalyst for Children at Risk of Poor Outcomes', https://www.adruk.org/learning-hub/user-groups-and-communities-1/adr-england-research-community-catalyst-children-at-risk-of-poor-outcomes/, accessed 4 Jun. 2025.

Anderson, M. et al. (2023) 'Children's Services Reform Research: Mapping Integration and Outcomes across Scotland: A Statistical Analysis', https://www.mendeley.com/reference-manager/reader/3da1d896-0ac4-3fc4-a315-a95a7a98b859/61319b56-38a0-0dc7-5659-97a608c04b1f, accessed 25 May 2025.

Baginsky, M., Moriarty, J., and Manthorpe, J. (2019) 'Signs of Safety: Lessons Learnt from Evaluations', *Journal of Children's Services*, 14: 107–23. https://doi.org/10.1108/JCS-11-2018-0028.

Bernheim, B. et al. (2025) 'Applying a Realist Evaluation to an Intervention in Children's Social Care: A Worked Example from the Safeguarding Family Group Conference Study', *Evaluation*, 31: 49–69. https://doi.org/10.1177/13563890241309643.

Bredewold, F., and Tonkens, E. (2021) 'Understanding Successes and Failures of Family Group Conferencing: An In-Depth Multiple Case Study', *The British Journal of Social Work*, 51: 2173–90. https://doi.org/10.1093/bjsw/bcab062.

Bunting, C. et al. (2026) 'Combining Target Trial Emulation and Qualitative Research to Understand the Effect of Health Visiting on Child Hospital Admissions in England', *American Journal of Epidemiology*, 195: 577–86. https://doi.org/10.1093/aje/kwaf207

Caffrey, L., and Browne, F. (2022) 'Understanding the Social Worker–Family Relationship through Self-Determination Theory: A Realist Synthesis of Signs of Safety', *Child & Family Social Work*, 27: 513–25. https://doi.org/10.1111/cfs.12903.

Caffrey, L., and Browne, F. (2023) 'The Challenge of Implementation in Complex, Adaptive Child Welfare Systems: A Realist Synthesis of Signs of Safety', *Children and Youth Services Review*, 150: 106992. https://doi.org/10.1016/j.childyouth.2023.106992.

Care Inquiry (2013) *Making not breaking - Building stable and loving homes for children in care*. London: Nuffield Foundation. https://www.nuffieldfoundation.org/sites/default/files/files/Care%20Inquiry%20-%20Full%20Report%20April%202013.pdf, accessed 25 May 2025.

Cavallaro, F. et al. (2024) 'Intensive Home Visiting for Adolescent Mothers in the Family Nurse Partnership in England 2010-2019: a Population-Based Data Linkage Cohort Study UsingPropensity Score Matching', *BMJ Public Health*, 2: e000514. https://doi.org/10.1136/bmjph-2023-000514

Children's Wellbeing and Schools Bill (2025), https://bills.parliament.uk/bills/3909, accessed 27 Feb. 2026.

Christie (2011) *Christie Commission on the future delivery of public services*. Edinburgh: Scottish Government. https://www.gov.scot/publications/commission-future-delivery-public-services/, accessed 27 Feb. 2026.

CLS. (2017) 'Scoping the Feasibility of a New Longitudinal Birth Cohort Study of Children Supported by the Child Welfare System: Recommendations for Next Steps', https://www.nuffieldfjo.org.uk/resource/born-into-care-newborns-in-care-proceedings-in-england-final-, accessed 25 May 2025.

Connelly, R., and Platt, L. (2014) 'Cohort profile: UK Millennium Cohort Study (MCS)', *International Journal of Epidemiology*, 43: 1719–25. https://doi.org/10.1093/ije/dyu001

Craig, P. et al. (2025) 'Using Natural Experiments to Evaluate Population Health and Health System Interventions: New Framework for Producers and Users of Evidence', *BMJ*, 388: e080505.https://doi.org/10.1136/bmj-2024-080505

Crawford, K. et al. (2025) 'Infant Mental Health Services for Birth and Foster Families of Maltreated Pre-School Children in Foster Care (BeST?): A Cluster-Randomized Phase 3 Clinical Effectiveness Trial', *Nature Medicine*, 31: 1617–25. https://doi.org/10.1038/s41591-025-03534-9

Crockett, R. et al. (2013). *Assessing the Early Impact of Multi-Agency Safeguarding Hubs (MASH) in London*. London: London Councils. https://basw.co.uk/sites/default/files/resources/basw_114425-8_0.pdf, accessed 25 May 2025.

Department for Education. (2025) *Evaluation strategy 2025–2029*. https://assets.publishing.service.gov.uk/media/680908f32a86d6dfb2b52782/evaluation-strategy-2025-to-2029.pdf, accessed 25 May 2025.

Drake, B., and Jonson-Reid, M. (1999) 'Some Thoughts on the Increasing Use of Administrative Data in Child Maltreatment Research', *Child Maltreatment*, 4: 308–15. https://doi.org/10.1177/1077559599004004004

Evans, R. E. et al. (2021) 'How Can We Adapt Complex Population Health Interventions for New Contexts? Progressing Debates and Research Priorities', *Journal of Epidemiology and Community Health*, 75: 40–5. https://doi.org/10.1136/jech-2020-214468

Family Rights Group. (2025) *What is a family group conference?* Family Rights Group. https://frg.org.uk/family-group-conferences/what-is-a-family-group-conference/?gad_source=1&gad_campaignid=22620091366&gbraid=0AAAADx98hBTn4OlfdMBqfEVGhYrEg9cP&gclid=CjwKCAjwl_XBBhAUEiwAWK2hzqkhvEnP6vJEfdRnH8N4fMv27_KHNhQWoiuFGyGI8nHA5CbFq0U0jxoCj6EQAvD_BwE, accessed 5 Dec. 2025.

Fitzsimons, A., and McCracken, K. (2020) Children's Social Care Innovation Programme Round 2 Final Report.

Forrester, D. (2017) Outcomes in children's social care. *Journal of Children's Services*, 12: 144–57. https://doi.org/10.1108/JCS-08-2017-0036

Gilbert, R. et al. *Evaluation of special educational needs and disability provision in English primary schools using administrative health and education data in the*

ECHILD database. MedRix, https://www.medrxiv.org/content/10.1101/2025.08.31.25334778v1, 2025, preprint: not peer reviewed.

Gillingham, P. (2018) 'Evaluation of Practice Frameworks for Social Work with Children and Families: Exploring the Challenges', *Journal of Public Child Welfare*, 12: 190–203. https://doi.org/10.1080/15548732.2017.1392391

Ginsburg, L. R. et al. (2021) 'Fidelity is Not Easy! Challenges and Guidelines for Assessing Fidelity in Complex Interventions', *Trials*, 22: 372. https://doi.org/10.1186/s13063-021-05322-5

Gubbels, J. et al. (2023) 'Long-Term Collateral Effects of Parent Programs on Child Maltreatment Proxies: Can Administrative Data Provide Useful Insights? ', *Children and Youth Services Review*, 155: 107260. https://doi.org/10.1016/j.childyouth.2023.107260

Harron, K. et al. (2017) 'A Guide to Evaluating Linkage Quality for the Analysis of Linked Data', *International Journal of Epidemiology*, 46: 1699–710. https://doi.org/10.1093/ije/dyx177

Hawe, P., Shiell, A., and Riley, T. (2004) 'Complex Interventions: How "out of control" Can a Randomised Controlled Trial Be?', *Bmj (Clinical Research ed.)*, 328: 1561–3. https://doi.org/10.1136/bmj.328.7455.1561

Heinrich, C., Maffioli, A., and Vázquez, G. (2010) A Primer for Applying Propensity-Score Matching Impact-Evaluation Guidelines. www.iadb.org, accessed 25 May 2025.

Hernán, M., and Robins, J. (2024) *Causal Inference: What If*. Boca Raton: Chapman & Hall/ CRC.

Hodges, H. R. et al. (2025) 'Why Social Work Needs to Embrace Data Linkage Research', *Journal of Children's Services*, 20: 1–9. https://doi.org/10.1108/JCS-12-2023-0063

Holmes, L. (2021) *Children's Social Care Cost Pressures and Variations in Unit Costs*. London: Department for Education.

Holmes, L., Neagu, M., Sanders-Ellis, D., and Harrison, N. (2020) *Lifelong Links Evaluation Report*. London: Department for Education.

Home Office. (2014) *Multi Agency Working and Information Sharing Project: Final Report*. https://assets.publishing.service.gov.uk/media/5a750ddb40f0b6360e472fa5/MASH.pdf

Hood, R. (2019) 'What to Measure in Child Protection?', *The British Journal of Social Work*, 49: 466–84. https://doi.org/10.1093/bjsw/bcy066

Hood, R. (2022) What linking children's social care and benefits data can tell us. *Blog post 25 January 2022*, Policy in Practice. https://policyinpractice.co.uk/what-linking-childrens-social-care-and-benefits-data-can-tell-us/, accessed 6 Dec. 2025.

Hood, R. et al. (2023) 'Categorising Demand for Child Welfare Services Using Latent Class Analysis: A Study of the National Data-sets on Children in Need in England', *The British Journal of Social Work*, 53: 3704–24. https://doi.org/10.1093/bjsw/bcad141

Hood, R. et al. (2024) 'Intervention Pathways following a Social Work Assessment: An Analysis of National Administrative Data for Children's Social Care in England', *The British Journal of Social Work*, 54: 2937–56. https://doi.org/10.1093/bjsw/bcae070

James, S., and Holmes, L. (2024) 'Evidence-Based Practice in Child and Family Social Work: Progress and Controversies', *The Routledge Handbook of Child and Family*

Social Work Research: Knowledge-Building, Application, and Impact, pp. 102–14. Milton Park: Taylor and Francis. https://doi.org/10.4324/9781003241492-8

Jay, M. A. et al. (2025) 'Estimated Cumulative Incidence of Intervention by Children's Social Care Services to Age 18: a Whole-of-England Administrative Data Cohort Study Using the Child in Need Census', *International Journal of Population Data Science*, 10: 2454. https://doi.org/10.23889/ijpds.v10i1.2454

La Valle, I. et al. (2019) How do we know if children's social care services make a difference? Development of an outcomes framework. https://www.nuffieldfoundation.org/project/measuring-outcomes-for-childrens-social-care-services, accessed 25 May 2025.

Jones, K. H. et al. (2018) 'The Good, the Bad, the Clunky: Improving the Use of Administrative Data for Research', *International Journal of Population Data Science*, 4: 587.

Méndez Pineda, R. (2026) 'Describing and Evaluating Multi-Agency Front-Door Services in England: A Proof of Concept Children's Social Care Retrospective Evaluation', PhD Thesis. UCL.

Méndez Pineda, R. et al. (2025) *Multi-Agency Safeguarding Hubs: Implementation and Process Evaluation Report*. https://foundations.org.uk/wp-content/uploads/2025/06/evaluation-of-multi-agency-safeguarding-hubs.pdf, accessed 6 Dec. 2025.

Netten, A. et al. (2012) 'Outcomes of Social Care for Adults: Developing a Preference-Weighted Measure', *Health Technology Assessment*, 16: 1. https://doi.org/10.3310/hta16160

Nurmatov, U. et al. (2020) Impact of shared decision-making family meetings on children's out-of-home care, family empowerment and satisfaction: a systematic review. [Project Report]. London: What Works Centre for Children's *Social Care*. Available at: https://foundations.org.uk/our-work/publications/impact-of-shared-decision-making-family-meetings-on-childrens-out-of-home-care-family-empowerment-and-satisfaction/, accessed 27 Feb. 2026.

Office for Health Improvement and Disparities. (2020) *Guidance: Patient-reported outcomes and experience study*. https://www.gov.uk/guidance/patient-reported-outcomes-and-experiences-study, accessed 25 May 2025.

Ottaway, H. et al. (2023) *Children's Services Reform Research: Learning and Implications for Scotland: Concluding Report*. https://www.celcis.org/application/files/6117/0245/9620/CSRR_Concluding_Report_-_CELCIS_-_13_December_2023.pdf, accessed 5 Jun. 2025.

Pawson, R., and Tilley, N. (1997) 'An Introduction to Scientific Realist Evaluation', *Evaluation for the 21st Century: A Handbook*, pp. 405–18. SAGE Publications, Inc. https://doi.org/10.4135/9781483348896.n29

Scottish Government (2016) Health and Social Care Delivery Plan. https://www.gov.scot/binaries/content/documents/govscot/publications/strategy-plan/2016/12/health-social-care-delivery-plan/documents/00511950-pdf/00511950-pdf/govscot%3Adocument/00511950.pdf, accessed 5 Jun. 2025.

Sebba, J. et al. (2017) Children's Social Care Innovation Programme Evaluation Report 58 Children's Social Care Innovation Programme Final evaluation report. https://assets.publishing.service.gov.uk/media/5a8202feed915d74e62354f0/Children_s_Social_Care_Innovation_Programme_-_Final_evaluation_report.pdf, accessed 24 May 2025.

Scourfield, J. et al. (2024) 'Family Group Conferencing for Children and Families: Evaluation of Implementation, Context and Effectiveness (Family VOICE). Study Protocol', *PLOS One*, 19: e0300834. https://doi.org/10.1371/journal.pone.0300834

Shadish, W. R., and Luellen, J. K. (2006) 'Quasi-Experimental Design', *Handbook of Complementary Methods in Education Research*, 3rd edn. Milton Park: Routledge.

Sheehan, L. et al. (2018) *Signs of Safety: Findings from a mixed-methods systematic review focussed on reducing the need for children to be in care*. https://whatworks-csc.org.uk/evidence/evidence-store/intervention/signs-of-safety/, accessed 18 May 2025.

Skivington, K. et al. (2021) 'A New Framework for Developing and Evaluating Complex Interventions: Update of Medical Research Council Guidance', *BMJ*, 374: n2061. https://doi.org/10.1136/bmj.n2061

Stabler, L. et al. (2019) Shared decision making: What is good practice in delivering meetings? Involving families meaningfully in decision-making to keep children safely at home: A rapid realist review. https://orca.cardiff.ac.uk/id/eprint/128345, accessed 27 Feb. 2025.

Stabler, L. et al. (2025) 'How Might Shared Decision-Making Meetings Reduce the Need for Children to Be in Care? A Rapid Realist Review', *Journal of Social Work*, 25: 102–24. https://doi.org/10.1177/14680173241258891

Suh, E., and Holmes, L. (2022) 'A Critical Review of Cost-Effectiveness Research in Children's Social Care: What Have We Learnt So Far?', *Social Policy & Administration*, 56: 742–56. https://doi.org/10.1111/spol.12795

Swansea University. (2024) *ESRC funding awarded to transform data collections for child welfare research - Population Data Science*. https://popdatasci.swan.ac.uk/esrc-funding-awarded-to-transform-data-collections-for-child-welfare-research/?utm_campaign=elcfs_update_202505&utm_medium=email&utm_source=newsletter&dm_i=6T0C, 1721O, 5XVW94,5LNJX, 1, accessed 25 May 2025.

Taylor, S. et al. (2024) 'Impact of Family Group Conference Referrals at Pre-Proceedings Stage on Child Outcomes: A Randomised Controlled Trial', *The British Journal of Social Work*, 54: 2358–77. https://doi.org/10.1093/bjsw/bcae035

Tucker, L. (2022) Quasi-Experimental Designs in Child Welfare Evaluations: Opportunities for Generating Rigorous Evidence. https://www.urban.org/sites/default/files/2022-10/Quasi-Experimental%20Designs%20in%20Child%20Welfare%20Evaluations.pdf, accessed 27 Feb. 2026.

UNCRC. (1989) *The United Nations Convention on the Rights of the Child*. https://www.unicef.org.uk/what-we-do/un-convention-child-rights/, accessed 25 May 2025.

Westlake, D. et al. (2024) *The Social Workers in Schools (SWIS) Trial September 2024 an evaluation of school-based social work: Follow-up report*. https://foundations.org.uk/our-work/publications/the-social-workers-in-schools-swis-trial-follow-up-report/, accessed 27 Feb. 2026.

Wood, S. et al. (2022) *A UK-wide survey of family group conference provision*. https://cascadewales.org/wp-content/uploads/sites/3/2022/12/CASCADE-Family-VOICE-survey-findings.pdf, accessed 24 May 2025.

Woodman, J. et al. (2025) *National Institute for Health Research Policy Research Programme Project: Evaluation of the 0–5 Public Health Investment in England: A Mixed Methods Study Integrating Analyses of National Linked Administrative Data With in-Depth Case Studies*. https://bit.ly/3JL9grh, accessed 6 Dec. 2025.

Woodman, J. et al. (2022) '*Study Protocol: A Mixed-Methods Study to Evaluate Which Health Visiting Models in England Are Most Promising for Mitigating the Harms of Adverse Childhood Experiences*', *BMJ Open*, 12: e066880. https://doi.org/10.1136/bmjopen-2022-066880 https://fundingawards.nihr.ac.uk/award/NIHR129901

Wright, J. et al. (2013) 'Cohort Profile: The Born in Bradford Multi-Ethnic Family Cohort Study', *International Journal of Epidemiology*, 42: 978–91. https://doi.org/10.1093/ije/dys112

Zhang, M. L. et al. (2020) 'Social Work Contact in a UK Cohort Study: Under-Reporting, Predictors of Contact and the Emotional and Behavioural Problems of Children', *Children and Youth Services Review*, 115: 105071. https://doi.org/10.1016/j.childyouth.2020.105071

British Journal of Social Work, 2026, 00, 1–21
https://doi.org/10.1093/bjsw/bcag017
Critical Commentary